From open data to reverse PSI

Poullet, Yves

Published in:
European Public Mosaic

Publication date:
2020

Document Version
Publisher's PDF, also known as Version of record

From *open data* to *reverse PSI* – A new European policy facing GDPR

Yves Poullet,
emeritus professor UNamur; associate professor UCLille, co-chairman of the Namur Digital Institute (NADI/ UNamur) (Belgium)

Abstract

In February 2020, the Commission published an expert report on the sharing of data collected by the private sector for the benefit of public authorities. Although the document is more of an invitation to the private sector than an obligation on the latter, it demonstrates a European will to strengthen the informational power of the State. Public authorities will be able to engineer big data, with the help of artificial intelligence, to better define state policies and their application. The article studies an essential regulatory facet, which must be taken into account when setting up this data sharing: compliance with the General Data Protection Regulation (GDPR), its principles, and the multiple obligations that the text imposes on the use of data provided by third parties, as well as the rights granted to citizens as data subjects. This brief overview suggests some difficulties in interpreting and applying the GDPR to these sharing operations.

1. The fight against COVID-19 as a starting point for our reflection

The fight against the pandemic currently being experienced worldwide testifies, in the eyes of the public, to a new phenomenon. Communication network operators in many countries are offering, under various formulas, to provide the public authorities with their subscribers' geolocation data. Whatever the reasons for this offer (whether self-interest or brand image) and the challenges posed by data protection requirements, the operators' proposal illustrates how sharing of information by the private sector for the benefit of the public authorities is useful for the common good. In the same context, the public authorities and citizens clearly need to be able to have information about the extent of the pandemic, region by region, published by Google and deduced from the crossed data collected through Google Now (a geolocation system linked to Android) and its search engine. We regret that this information was not shared and refined directly in consultation with the public health authorities.

Our purpose is not to go into what data sharing should be or should have been in the context of the fight against COVID-19. Apart from this example, we wonder whether there is a need for a coherent

We wonder whether there is a need for a coherent data protection policy on the sharing of data collected by the private sector for the benefit of public authorities in the public interest

data protection policy on the sharing of data collected by the private sector for the benefit of public authorities in the public interest. Our reflections are inspired – even if we do not entirely agree with the report's analysis and conclusions – by the recent documents produced by the EU Commission on B2G data sharing: in particular, the report on the topic prepared by the high-level expert group (HLEG) on business-to-government (B2G) data sharing.[1]

1 *Towards a European strategy on business-to-government data sharing for the public interest: Final report prepared by the High-Level Expert Group on Business-to-Government Data Sharing* (2020). The EU Commission presented this document on February 19 as part of its data policy strategy.

2. A paradigm shift in a new technological context

The public sector has traditionally been a source of data for the private sector. This conception is rooted in freedom of expression, which requires citizens and businesses to be able to have access to the information on which the public authorities' actions are based. This transparency of public powers has given rise to the demand for a proactive state policy to make data available to companies, allowing them to be reused and fully exploited. The recent Open Data Directive[2] has further expanded the state's duty to the whole of the public sector, prescribing the need to use an open data format, which allows maximum reuse.[3] The shift we are describing reverses this unilateral direction of flow. It is now the public authorities that have become the recipients of flows from the private sector, hence the name "reverse PSI Policy".[4] The concept of *reverse PSI* refers to the

2 **Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the reuse of public-sector information** (OJ L 172, 26.6.2019, pp. 56-83), also known as the *Open Data Directive*, aims to make more public-sector information available and reusable.

3 According to the famous FAIR (findability, accessibility, interoperability, reusability) principles (regarding the FAIR principles, see the ***FAIR Guiding Principles for scientific data management and stewardship***). These principles were published in *Scientific Data*. The authors intended to provide guidelines to improve the capacity of computational systems to find, access, interoperate, and reuse data with little or no human intervention. The 2017 Tallinn EU Ministerial Declaration on e-Government calls on governments to "increase the findability, quality and technical accessibility of data in key base registers" and suggests extending these principles to the private sector in order to facilitate sectoral and inter-sectoral data sharing.

4 Comission Staff Working Document. ***Guidance on sharing private sector data in the European data economy***. Accompanying the document *Communication from the Commission to the European Parliament, the Council, the European economic and social Committee and the Committee of the Regions. "Towards a common European data space"*.

Public Sector Information (PSI) Directive, which creates a right to re-use all public documents (data) held by Member States' public sector bodies. Reversing the concept of the PSI Directive would entail access by public sector bodies to re-use privately held data.

It is now the public authorities that have become the recipients of flows from the private sector

Why is the Commission seeking to bring in this new policy? The reason is twofold. Firstly, the public authorities wish to be able to use artificial intelligence (AI) tools both to define their policies and also to ensure their effectiveness. Creating urban traffic plans requires precise traffic data. If the public authorities had to put technical means in place to measure it, collection would cost a hundred times more than access to data that transport and navigation software companies or collaborative platforms such as UBER collect as part of the services they offer. Bringing supply and demand for employment together in the best possible way may require our public employment assistance agencies to have socio-economic data from companies or professional associations. This would guide training efforts and define the desired employee profiles. In short, AI requires the existence of a sufficiently rich and numerous data set so that complex *machine learning* algorithms can identify statistically significant correlations. The second reason, in the case of a public authority, is that the source of the data collected was traditionally internal to the public authority. It was only rarely external and even then limited to very specific files. As the examples show, the need for the

administration to use the most adequate technologies to define and achieve the common good (and AI can do this if certain conditions are met) justifies access by the administration to data collected only by the private sector. In addition, if public authorities do not have such access, they find themselves in a position of inferiority and at the mercy of private operators that have more accurate, available and up-to-date information. The case of Google and the spread of the pandemic is a perfect example of this (see point 1 above).

3. A prerequisite: the definition of private sector data use

We fully support the principle of reverse PSI and the use of artificial intelligence. It is our view that this should take place with the necessary strengthening of the State at a time of increasing privatization of information and therefore of power. However, at the same time, we would like to recall some basic rules of our data protection legislation that the State cannot dispense with. The first rule concerns the very legality of the public authorities using private sector data. For various reasons, some more legitimate (corporate branding) than others (collusion with the administration, obtaining a regulatory advantage), private companies may wish to "offer" their data or to create joint ventures with public authorities. We are of the opinion that a regulatory framework should apply to all of these proposals. As stated in the HLEG in the B2G data sharing report commissioned by the EU

The need for the administration to use the most adequate technologies to define and achieve the common good, like AI, justifies access by the administration to data collected only by the private sector

It is necessary to strengthen the State at a time of increasing privatization of information and therefore of power

Commission, one can imagine citizens themselves, with their consent, wishing to contribute to the public interest by offering their data. Nevertheless, such flows or sharing can only be justified under Article 6 of the GDPR in the context of a specific public interest purpose, previously set within the context of a "law" in the broad sense of the term, which is transparent, proportionate and necessary in a democratic state. Therefore, the public authority must define the purposes of the B2G data sharing.[5]

The public authority must thus precisely define the purposes pursued: for instance, assistance to the unemployed, urban planning, the definition of a transport policy, medical research, etc. We also cannot exclude control of tax or benefit fraud.[6] It is important for the authority to be able to clearly demonstrate that the public interest benefits are greater than the disadvantages for citizens or economic partners. There can be no question of creating big data that can be used for all "useful" public interest purposes; only those that come within the framework of explicit legal purposes compatible with the GDPR. With the exception of statistical offices, whose operation is subject to strict confidentiality rules, there can be no

5 The French Act (***Loi du 7 octobre 2016 pour une république numérique***, Art.17 and following) is very interesting regarding this point. It allows the public sector to access data held by the private sector in certain contexts and obtain them from certain actors when it is in the public interest.

6 The HELG report excludes them since it would give B2G data sharing a poor image. Another purpose is excluded: use of private sector data for commercial purposes. It is quite clear that commercialization would go beyond the role of public administration and distort the competitive private market.

The French Digital Republic Act was passed in 2016 and it was pioneering as it allows the public sector to access data held by the private sector when it is in the public interest.

question that the decision should depend only on negotiations between supplier companies and the administration.[7] If the law provides for the possibility of citizens providing data concerning themselves collected or processed by the private sector, that consent can only be given within the framework of the legal purposes pursued.

It is in light of such purposes that the extent and quality of the data requested from the private sector should be assessed. Such purposes will determine the extent and quality of the data requested from the private sector, the degree to which data is provided raw or aggregated, and the frequency of updates and access. However, this principle of

7 This does not exclude discussing the forms of the data flow with the companies (format, compensation for the costs they have incurred, etc.).

minimization, called for by the group of experts[8] poses difficulties when it comes to the public authority setting up artificial intelligence systems, especially those that are unsupervised and involve deep learning. These systems are characterized by the fact that the system provides significant correlations without knowing what data will be useful at the outset. Notwithstanding this precaution, we should take into account, firstly, that the administration already has certain data and that there can be no question of duplicating sources. Secondly, we should consider sorting useful data after some testing and, perhaps, experimentation.

The principle of minimization poses difficulties when it comes to the public authority setting up AI systems, especially those that are unsupervised and involve deep learning

4. The need for PIA of data sharing operations

The sharing of data between the private sector and public authorities requires an assessment of the risks involved. Article 35 of the GDPR requires such an assessment in the case of so-called high-risk processing[9] and we believe that most of the processing generated in the context of these "private-public" flows must be considered as such. The main reasons for this are that the processing thus generated will target the general population or at least a part of the population, whether it is processing to help define public actions or

8 "The requested private-sector data should be necessary, relevant and proportionate in terms of detail (e.g. type of data, granularity, quantity, frequency of access) with regard to the intended public interest pursued" (Report, *op.cit.*, p. 80).

9 Regarding the concept of *high-risk* and the interpretation to be given to Article 35 of the GDPR, see EDPB, ***Guidelines on data protection impact assessments high risk processing*** (2018).

to implement regulations. The processing will thus have a significant impact on citizens. In addition to these considerations and the risks already covered by data legislation, there are risks of discrimination or social injustice linked to programming errors and especially to biases. These may be due to the quality or completeness of the programs used or whether the data are provided raw or preprocessed downstream.

We believe that this assessment must be imposed and be entrusted to a multidisciplinary and multistakeholder body in close cooperation with the data protection, civil liberties and equal opportunities authorities. Without interfering with the competencies of those other authorities, that body will have to coordinate the opinions of these different authorities and examine both the technical aspects, the objectives pursued and the means to handle them (including the effective need to resort to private-sector data). Such a body, called *Data Ethics*, has been created in the United Kingdom, Denmark and Germany.[10] It is interesting to note that in its "Data Strategy" document dated February 19, the Commission also calls for the creation

The sharing of data between the private sector and public authorities requires an assessment of the risks involved, like discrimination or social injustice

10 In the UK, the **National Statistician's Data Ethics Advisory Committee** (NSDEC) has been established to advise the National Statistician that the access, use and sharing of public data, for research and statistical purposes, is ethical and for the public good. In Denmark, the Danish Council of Ethics was set up in 1988 (see *The History and spheres of work of the Council* ***https://www.etiskraad.dk/english***). In Germany, the *Datenethikkommission* was established in 2018. The German commission issued a very important **report** on data ethics covering both data and algorithmic systems.

The UK, Denmark and Germany have multistakeholder bodies in charge of *data ethics*. These bodies should work in close cooperation with authorities in charge of data protection, civil liberties and equal oportunities

of such a body.[11] That body will submit its report to the legislative authority, which has to balance the interests of all stakeholders: public authorities, private companies, associations and citizens, taking into account the risks mentioned above. The report will be published and the assessment should be repeated at regular intervals, as artificial intelligence systems evolve, based on the data received and new sources that may be used.

5. What about data subjects' rights?

Regarding data subjects' rights concerning the processing created by B2G sharing data, Article 23 stipulates: "Union or Member State law to which the data controller or processor is subject may restrict by way of a legislative measure the scope of the obligations and rights provided for in Articles 12 to 22 and Article 34, as well as Article 5 in so far as its provisions correspond to the rights and obligations provided for in Articles 12 to 22, when such a restriction respects the essence of the fundamental rights and freedoms and is a necessary and proportionate measure in a democratic society to safeguard [...] e) other important objectives of general public

11 "Given already existing structures such as in finance, pharmaceuticals, aviation, medical devices, consumer protection, data protection, the proposed governance structure should not duplicate existing functions. It should instead establish close links with other EU and national competent authorities in the various sectors to complement existing expertise and help existing authorities in monitoring and the oversight of the activities of economic operators involving AI systems and AI-enabled products and services" (p. 26 of EU Commission ***White paper on artificial intelligence - A European approach to excellence and trust***, 2020, COM, 2020, 65 final).

interest of the Union or of a Member State, in particular an important economic or financial interest of the Union or of a Member State, including monetary, budgetary and taxation a matters, public health and social security". This broad exemption could apply with regard to the sharing of data between private and public authorities but must be duly justified, with regard to each of the different obligations (obligation to provide information, including in the event of security breaches, obligation to respond with regard to the data subject's right of access, etc.) and with regard to each of the various rights attributed to the data subject by the GDPR (right of access, right to de-referencing, right to object, etc.).

In addition, as requested by the HLGE (2020 report, p. 85), it is necessary for the sharing operation to be subject to great transparency: "Business-to-government data collaborations should be transparent about the parties to the collaboration and their objectives. Where possible, public bodies should also be transparent on the data that has been used and the algorithms applied, as well as on the results of the collaboration, including the relation to subsequent decision-making and the impact on individuals. Moreover, public bodies should ensure ex post transparency to the private companies and civil-society organisations on which particular public interest has been advanced with the use of their data and how, and cases where the data has not been used. Good practices should be made publicly available".

Business-to-government data collaborations should be transparent about the parties to the collaboration and their objectives

6. Withdrawal of consent and the right not to be subject to automated decisions

The long list of possible exemptions does not cover two hypotheses: the first is withdrawal of consent (GDPR, Art. 7. 3). Consent may be the basis for data processing by the private sector; according to the idea expressed in the HLGE report of consent qualified as altruistic, it can also be expressed in a way that is certainly not a sufficient basis (see point 3 above) for the transfer of private sector data to the public administration. With regard to the second hypothesis, one can imagine patients suffering from a disease expressing their agreement to doctors, pharmaceutical companies and hospitals transferring their data to a data bank maintained by the ministry of public health or car drivers asking the car maker to share their data with the Ministry of Transport. Withdrawal must be possible in both cases. They will naturally contact the private company, unless otherwise indicated. It will therefore be important to allow for withdrawal to be reported to and acted upon by the administration, which

in practice may prove difficult or even impossible.

The second point is more delicate: it concerns Art. 22 of the GDPR, which excludes that a decision may be taken solely on the basis of an automated system. However, the same article introduces exceptions, in particular in case of a legal authorization, subject to "suitable measures to safeguard the data subject's rights and freedoms and legitimate interests" (Art. 22.2.*b*). This exception obviously applies to any automated administrative decision beyond those based on data shared with the private sector. As regards all these automated decisions, we suggest that the French legislation on the relationship between the citizen and the administration[12] be followed. Firstly, it creates an obligation, insofar as possible, to reveal the source code as a communicable administrative document. Secondly, "an individual decision taken on the basis of algorithmic processing shall include an explicit notification informing the person concerned. The rules defining this processing and the main characteristics of its implementation shall be communicated by the administration to the interested party if he or she so requests". This means that, subsequently, information must be spontaneously provided by the administration, in an intelligible form, concerning the degree

As regards automated decisions, we suggest that the French legislation on the relationship between the citizen and the administration be followed

12 See Art. 2 of *Loi pour une République numérique* and Art. L 300-2 of *Code des relations du public avec l'administration*. See also French ***Décret nº 2017-330 du 14 mars 2017 relatif aux droits des personnes faisant l'objet de décisions individuelles prises sur le fondement d'un traitement algorithmique*** (JORF no. 64, March 16, 2017).

We are gradually shifting from an obligation to provide information to an obligation to provide an explanation

and mode of contribution of algorithmic processing to the decision-making, the data (personal or not) processed, their sources (private or public), and the processing parameters. If applicable and where appropriate, the information must also include the weighting of the parameters applied to the situation of the person concerned, as well as the operations performed by the processing. The main idea behind these provisions is to offer the data subject greater monitoring of the decision and to effectively empower him or her to contest it. We are gradually shifting from an obligation to provide information to an obligation to provide an explanation. Furthermore, just like other EU documents,[13] French legislation requires oversight of the algorithmic system in operation. Administrative decisions therefore cannot be based solely on a learning algorithm in which no control or validation by a human being is required.

7. Conclusions

The recent European approach has been expressed in a very measured way. This approach is based more on a desire to convince the private sector of the interest in collaborating with the public sector than on a real proactive policy for the development of modern and rich information systems within the administration to support the public authorities in directing and achieving

13 In particular, see the HLGE on artificial intelligence ***Ethics guidelines on a trustworthy AI*** (2019) and more recently the EU *White paper on artificial intelligence - A European approach to excellence and trust* (2020) (COM, 2020, 65 final).

the public interest. Notwithstanding this downside, the HLGE report reveals the increasingly necessary consideration of the public interest and the need to strengthen the informational power of the State compared to that of the private sector. In relation to this desire, the GDPR appears (and this is to be welcomed) to be establishing guidelines for such "data sharing" between the private and public sectors. It emphasizes the need for a legal basis or the performance of public missions, the requirement to inform citizens of the establishment of information systems, and the need for human validation and explanation of decisions made on an algorithmic basis. At the same time, the document suggests the need to go beyond the purely individualistic concerns that underpin the provisions of the GDPR. The public interest is a particular focus and citizens are even asked to contribute through "altruistic" consent and not merely defend their private interests. The evaluation of this interest and "proportionate" methods to achieve it should be the subjects of a multidisciplinary and multi-party public debate, which the privacy impact assessment of the GDPR has certainly commenced but which must finally take into account other more collective dimensions, in particular related to human dignity and social justice.[14] ▮

The HLGE report reveals the increasingly necessary consideration of the public interest and the need to strengthen the informational power of the State compared to that of the private sector

14 See Frenay, Benoît, & Poullet, Yves. (2019). ***Profiling and Convention 108+: Report on developments after the adoption of Recommendation (2010)13 on profiling***. Consultative Committee of the Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data. The report analyses various recent Council of Europe publications mentioning the need for enlargement of the societal concerns to be taken into account regarding machine learning technologies.